

A corpus-based analysis of indefinite article use in London English

Costas Gabrielatos & Eivind Torgersen
(Lancaster University)

Part of the project:

An analysis of spoken London English using corpus tools

PI: Eivind Torgersen
CI: Paul Kerswill
RA: Costas Gabrielatos
Consultant: Sebastian Hoffmann



Overview

- Link to previous and current projects
- Theoretical considerations
- Data and methodology
- Analysis and conclusions (so far)

*Linguistic innovators:
The English of adolescents in London (2004–
7)*

*Multicultural London English:
The emergence, acquisition and diffusion
of a new variety (2007–10)*

Investigators:

Paul Kerswill (Lancaster University)

Jenny Cheshire (Queen Mary, University of London)

Research Associates:

Sue Fox, Arfaan Khan (Queen Mary, University of London)

Eivind Torgersen (Lancaster University)

Funded by the Economic and Social Research Council

<http://www.lancs.ac.uk/fss/projects/linguistics/innovators/>

<http://www.lancs.ac.uk/fss/projects/linguistics/multicultural/>

E · S · R · C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL



Why study London English?

- London as the centre of linguistic innovation in British English
 - Diffusion of linguistic features from inner to outer London and beyond
- London as a multicultural city
 - High level of dialect and language contact

Theoretical background

- Variation in the use of indefinite article in ...
 - British English dialects
 - Child language
 - Contact varieties
 - L2 varieties

Lack of *a/an* alternation in British English dialects

- Wright (1905:71):
‘very few dialects follow the rule of the literary language according to which *an* is used before a vowel’.
- Shorrocks (1999:45) on the Bolton dialect:
‘the indefinite article is pronounced /ə/ or /ən/ before vowels and diphthongs’.
- Wagner (2008:418) on dialects in Southwest England:
‘the indefinite article often occurs as *a* before vowels’.

Survey of English Dialects

- One response: ‘*a/an* April fool’
- ‘*a* April fool’ is *more frequent* than ‘*an* April fool’ in localities in:
 - The East Midland counties;
 - Somerset (the rest of the Southern counties have not recorded the article or it was undetected);
 - The West Midland counties;
 - The Northern counties
- So basically everywhere!

Lack of *a/an* alternation in child language

- Newton and Wells (1999)
- Percentage of 'adult' examples increased from just below 30 at the age of 3 to 80 at the age of 7
- The development follows very much that of the definite article and other morphology

***a/an* in African American Vernacular English (AAVE)**

- Labov (1972) *Language in the inner city*:
- ‘He is a expert’
- ‘Bell mother got a old beat-up boot’

***a/an* in L2 varieties of English**

- Seliger (1979)
- Studied *a/an* alternation among young L2 speakers of English (younger than 13 and adult learners)
- Preference for *a* in both vowel and consonant environments, but used *an* more frequently with increased age

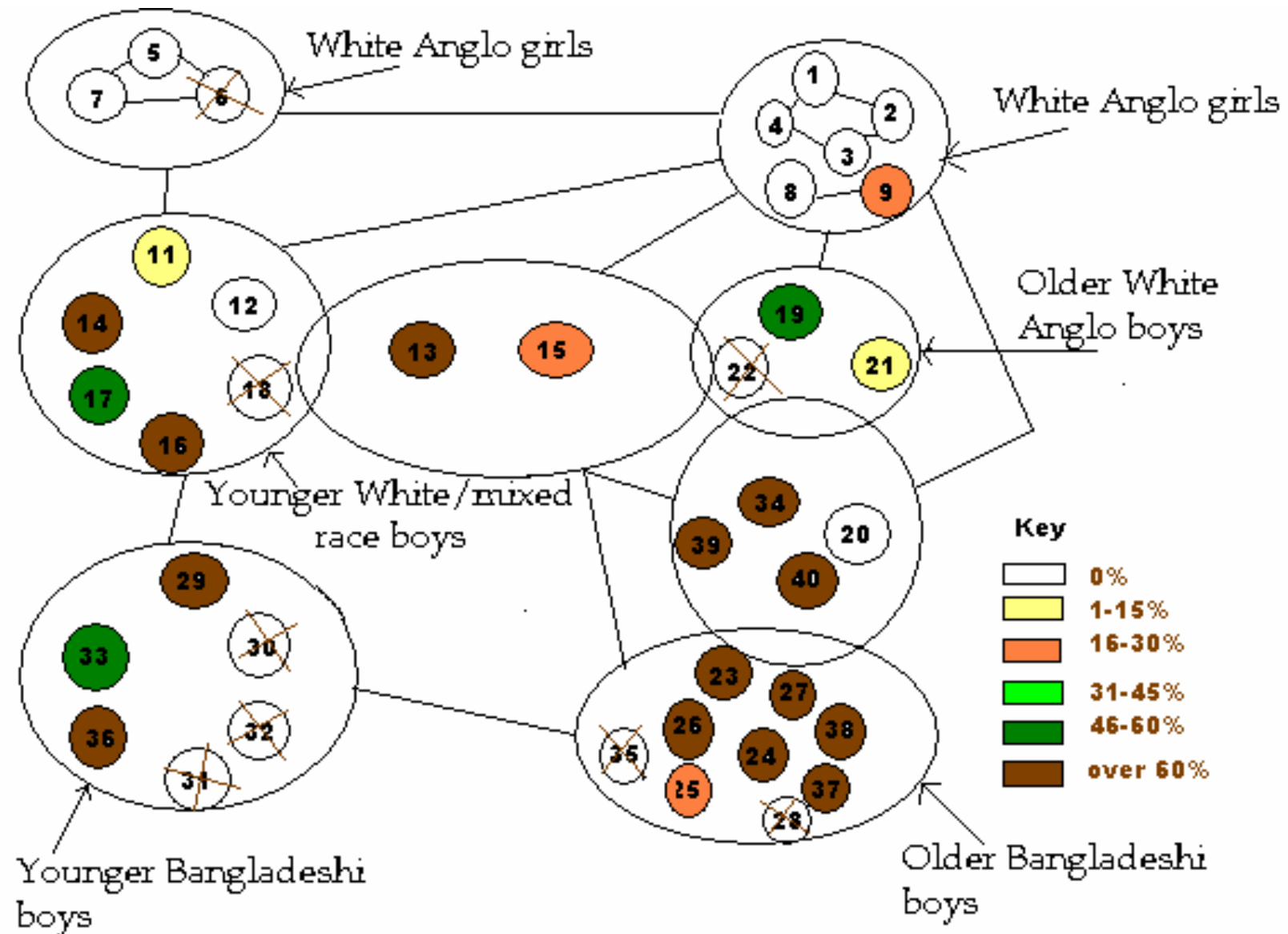
Tower Hamlets (Fox, 2007) (1)

- Tower Hamlets (London borough just South of Hackney)
- 39 speakers
- Age: 12-17
- 9 girls (all of White British origin)
- 30 boys (11 White British, 17 Bangladeshi, 2 mixed)
- Small sample: 94 indefinite article tokens in pre-vocalic position
- Before vowels:
 - *an*: 65%
 - *a*: 35%

Tower Hamlets (Fox, 2007) (2)

- Use of *a*
 - Bangladeshi boys: 75%
 - Mixed race White/Afro-Caribbean boys: 35%
 - White boys: 15%
 - Boys: 44%
 - Girls: 5%

Friendship groups in Tower Hamlets



Research questions and hypotheses

- What variables are good predictors of use?
- Effect of phonological context?
- Non-Anglo boys in Hackney will be the highest users.
- A person with a multicultural friendship group will use *a* before vowels more often.

Data

- Corpus of approx. 1.4 million words, transcribed orthographically.
- 117 speakers:
 - L1 = English
 - Some bilingual
- 1,241 tokens
- Five variables:
 - Sex: female (53), male (64)
 - Age: old=70+ (18), young=16-18 (99)
 - Ethnicity: Anglo (77), non-Anglo (40)
 - Residence: Hackney (58), Havering (59)
 - Network (only 'young'): 1 (16), 2 (9), 3 (19), 4 (25), 5 (15)

Methodology (1)

- Concordance analysis and annotation of *a* and *an* (sorted at R1, R2, R3) to identify (and **count**):
 - instances of ‘indefinite article + adjective/noun’ (continuous or discontinuous)
 - speaker identity of ‘*a* + vowel’ users

N	Concordance	Set	Tag
88	off Sue: so that's been abused is it? Grace: yeah /([Maria: it is])/ because . a a girl was on a train . . because her bluetooth was on . and some man		R
89	the wrong one Hadiya: it's cool Sue: was it? Hadiya: yeah Sue: is that a a girl's school or mixed? Hadiya: it's a mixed school is a near . like in		R
90	and then pearl harbour in i think november or december '41 so they were a a glimmer of hope . especially in '42 or '43 . when the germans were		R
91	it's great but it's a [long flight] Monty: [and i i] want to go to india as well . a . a . goa my daughter went there my middle daughter went there . they		R
92	movement [mm] erm . er in the trade union movement because i i'm a . a gold member . retired of the um now the gmb er municipal . workers		R
93	i mean if they'd had done it on the television you'd have thought it was a . a good comedy sketch . apparently if you go there . and you pay with		R
94	to sleep something and i always finish up by saying thank you for being a . a good audience and those who went to sleep thank you for not snoring		R
95	Ollie: oh yeah yeah yeah Sue: what about you? who . you gave me a a group of people . who would you say was your closest friend? ...		R
96	mm Talal: like . it's just . the Rom the Romford people they they got like a . a high tone innit comes out of their mouth like .. like they they bring		R
97	quite good and they've tried to do things and they've tried to open ?? of a a history club there you know and . i'm there this afternoon i go along		R
98	you span .. well .. not the entire century like the queen mother did but a a huge chunk of it [Mrs Knight: mm] and the changes that's been [Mrs		R
99	[Sue: mm] and erm when I used to steam er blanche in boiling water in a a I had a special saucepan with like a mesh thing which used to be		U
100	know years ago you know bits of schrapnel you know so it was it was a a i suppose to children it was er er can i say an exciting time you know		U
101	the teachers used to go to him and er /used to but he er his . he . suffered a . a . illness i don't know what it was and his boy took it up but then he		R
102	right so she didn't move very far either Arthur: no that's right no no . cos a . a . in those days there was a tendency to live near your fam. i.i. had what		U
103	on the staff [Sue: mm] so he got the pension and [Sue: oh I see] a good a a index linked pension which proved to be very beneficial [Sue: mm] so		R
104	..just bef just before he he passed away we went to cuba together .. a a a it it was fun [Sue: yeah] it's nice I mean it's ss so it's like the .. so m		U
105	. telly on the er radio . on sundays they . er there's a chap there he gives a . a jazz concert and he gets all different places as what you know .		R
106	of achievement and that and then /([Sue: yeah])/ me and my friends got a a jeep .. Bisa: really? we got in a limo Isabella: I think we fitted fourteen		R
107	diseases in supply there he's he's been there a number of years he had a a job in er er . in pensions or something of that nature but . something he		R
108	joint income so i mean it's an impossibility to buy anything e. even getting a a joint income a joint mortgage [yeah] because if you i. i. if your got a		R
109	er er municipal properties to housing associations . and they're making a a killing aren't they i mean you got the peabody you got the guiness you		R
110	a problem here in a minute like can you try and help man out yeh give me a . a knife or something yeh cos i'm gonna need to stab through" i mean		R
111	in the schools . bad spelling . bad writing [mm] . sometimes you get a a letter from somebody a younger person and you can't read it . even		R
112	Sue: oh right did you get those? Adam: y- actually I wan- I wanted to do a A level but when I was at the exams they told me to go straight to the .		A
113	they're safe and the other people are not . whereas I've grown up with a . a like more of white society . and more of a like and like some Asians		R
114	mind you're saying you need more .. but like if you . like say you had a a line one night .. and . you think "I want more and more" but the next		R
115	you to make things but . it's not got a label on it you know it's got to be a a little tick on it [Shirley: designer] or something . even even the young		R
116	where did you go Edie: i went to eastbourne .. and er . i finished up in a a little village called willingdon that's about . five miles from the coast of		R
117	running and then he start . popping it after him . yeah but . there was a . a /lot of people running/ Sue: this was one of your erm lot was it? Brian: R		R

Annotation examples

Category	Example
<i>a</i> is not an article	▪ <i>to have a a at least</i>
Repetition	▪ <i>a a book</i>
Self-correction	▪ <i>have you got a an extended family?</i> ▪ <i>it would go like a like two years</i>
Hesitation / fillers	▪ <i>a erm</i>
Unfinished utterances	▪ <i>... you know what I mean it's a. you might live in Clapton and you ...</i>
Backchannel interruption	▪ <i>Sue: mm what's / [Fatima: I liked my a] / your favourite subject at school</i>
Humorous or metalinguistic uses of 'a + vowel'	▪ <i>a orse</i> ▪ <i>... you can't say a aeroplane you have to say an aeroplane</i>

Methodology (2)

- Tabulation of speaker information:
 - user / non-user
 - tokens and types
 - variable values
- Manual calculation of frequencies and cross-tabulations.

Methodology (3)

- Metrics
 - tokens / speakers (*density*)
 - users / speakers (*spread*)
 - Rough indicator: $density \times spread$ (*preference*)

	Speakers	Users	Tokens	<i>Density</i>	<i>Spread</i>
Male	64	37	119	1.86	57.8%
Female	53	30	63	1.18	56.6%
Total	117	67	182	1.56	57.3%

Low predictive utility

- **Phonological context**

- No pattern emerged from the examination of use with *a* and *an*.
- Analysis was in terms of phonemes **not actual phones**.

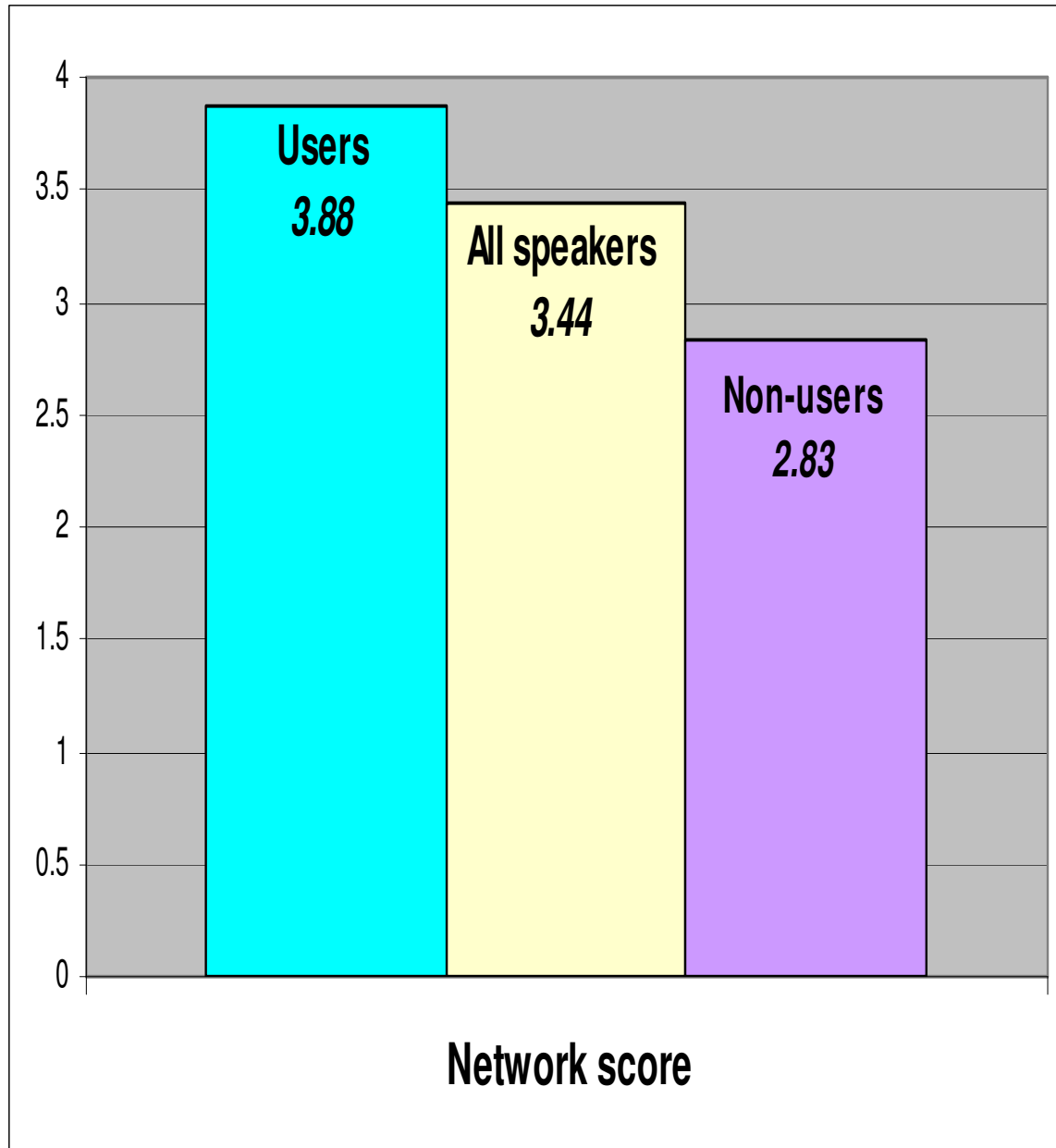
- **Network score**

- Correlation of network scores to usage preference symptomatic of the very strong correlation between usage and ethnicity.
- High network multi-ethnicity not a good predictor overall.

However

- Network score good predictor of use among Anglo speakers.

Network score (1)

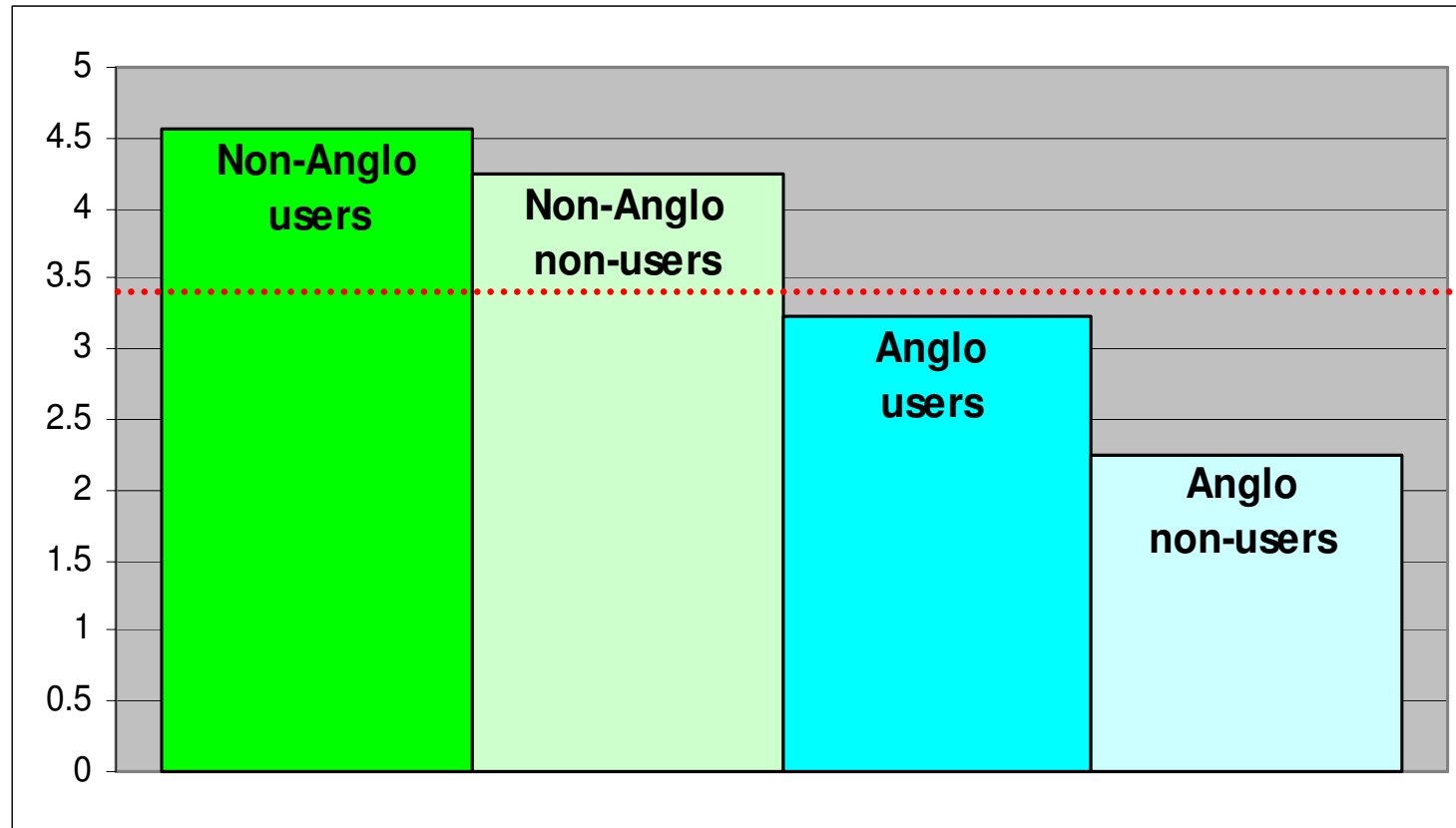


Average network score of users almost 40% higher than that of non-users.

But

Difference of the users' network score from average is about 50% smaller than that of non-users.

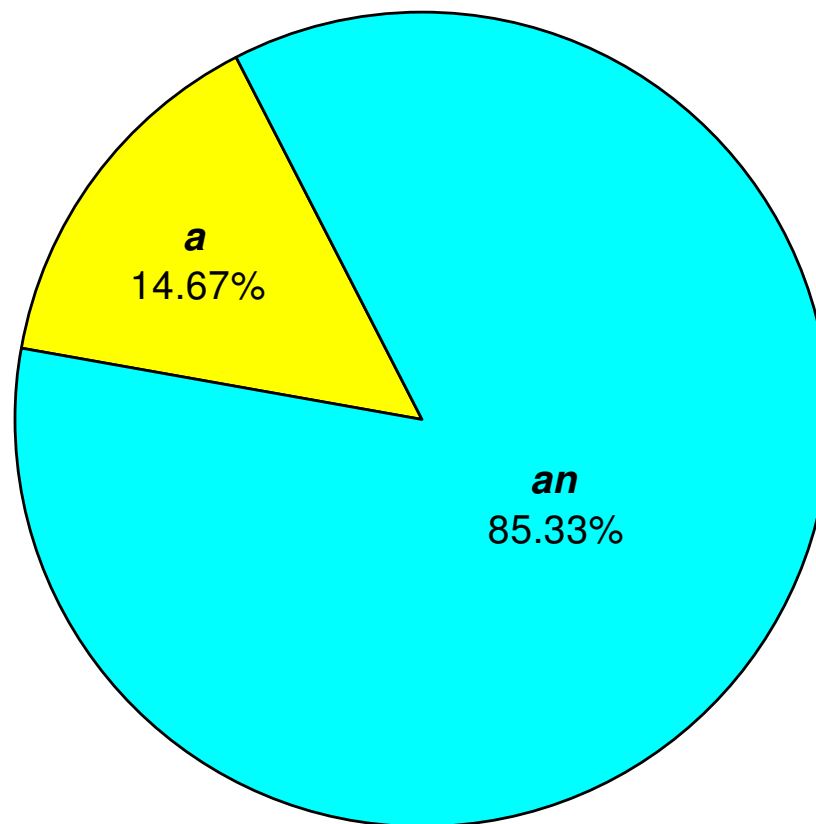
Network score (2)



The seeming correlation between network scores and use is probably due to the correlation between network score and ethnicity (AA=2.75; NA=4.48).

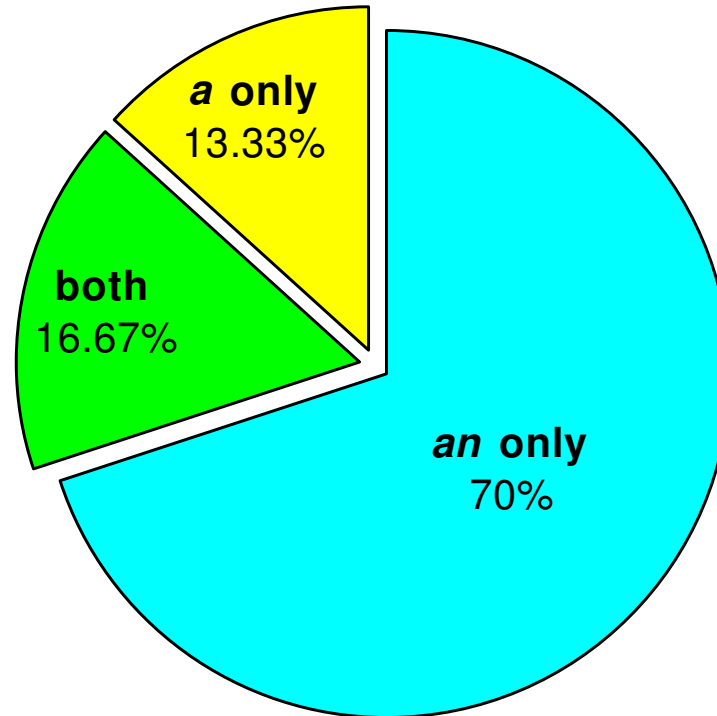
Indefinite article use: tokens

	Raw freq.
<i>a</i> + vowel	182
<i>an</i> + vowel	1,059
Total	1,241



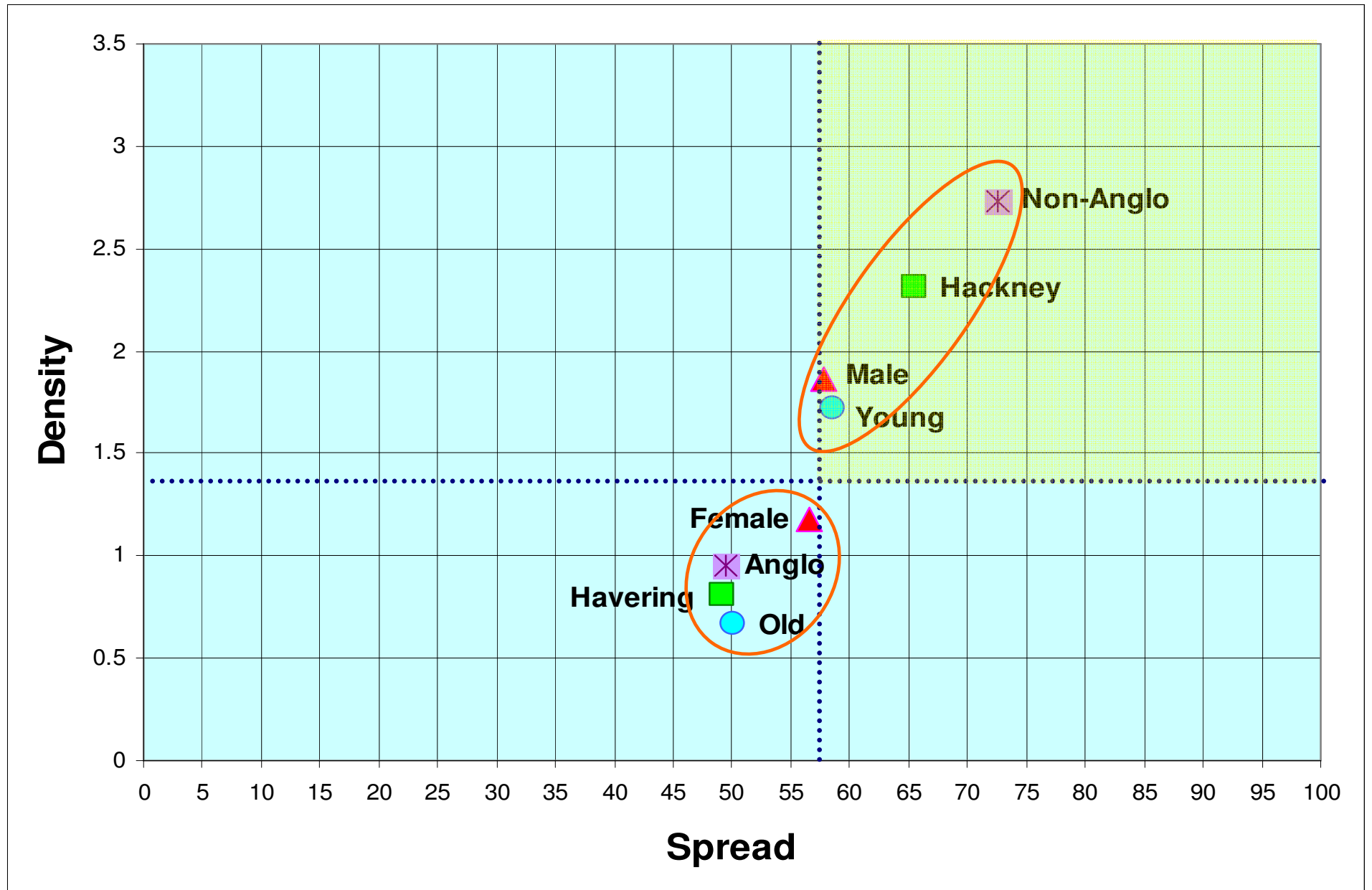
Indefinite article use: types

	Raw freq.
<i>a</i> only	48
<i>an</i> only	252
Both	60
Total	360



Use in terms of individual variables

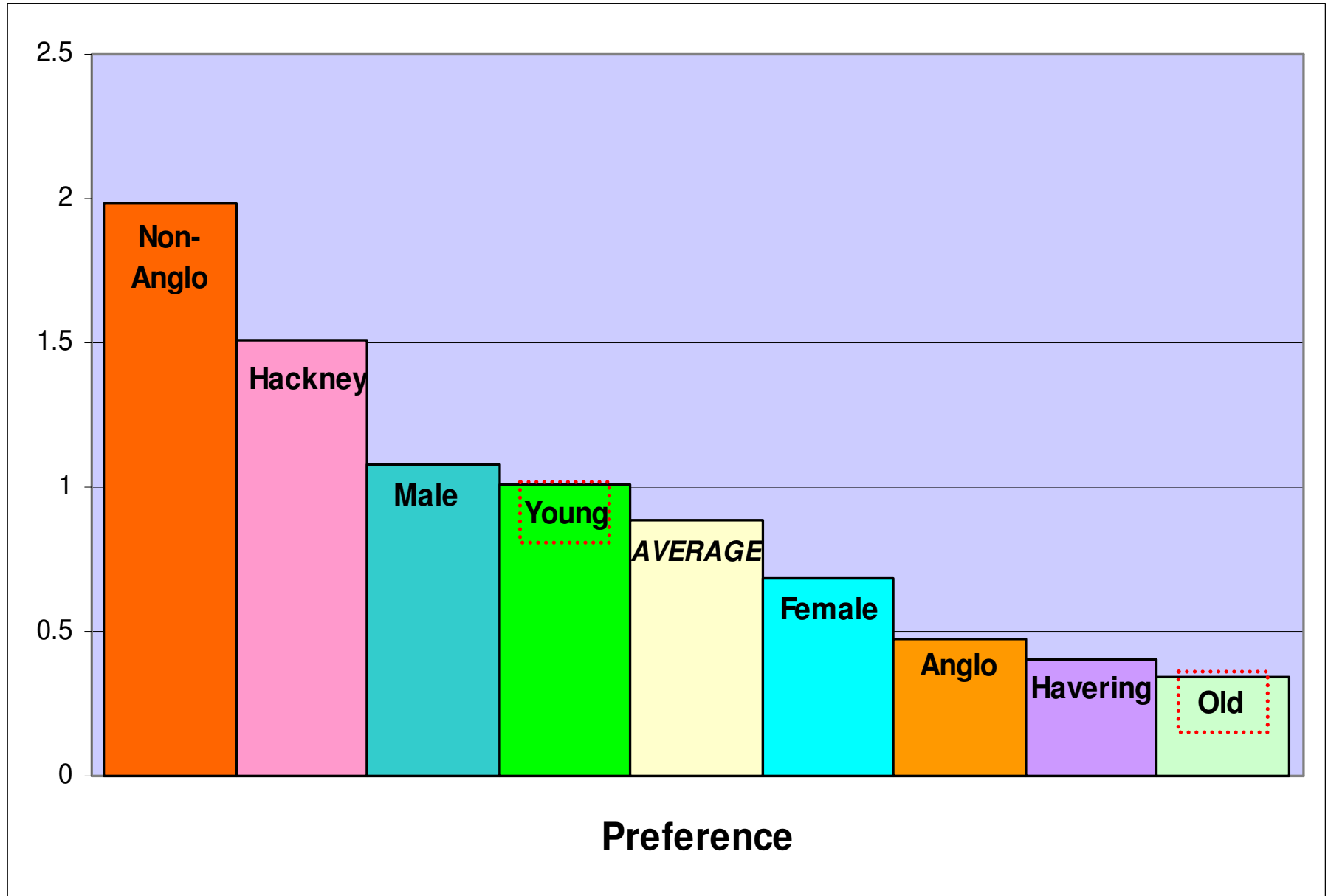
Interaction of density and spread



Comments

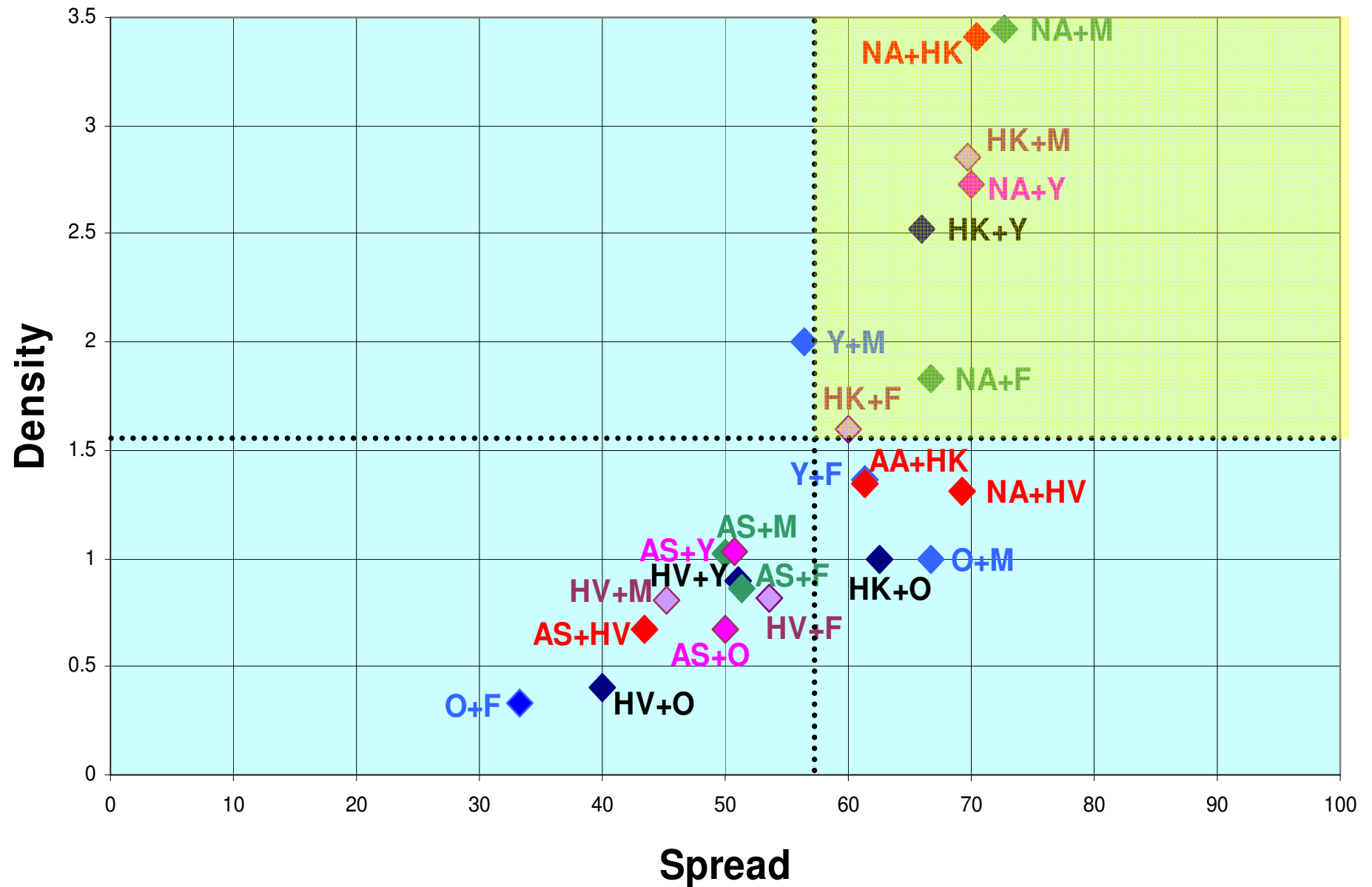
- Unlikely that clusters have formed by chance.
- Typical user profile:
[NON-ANGLO] [YOUNG] [MALE] [HACKNEY RESIDENT]
- [NON-ANGLO] and [HACKNEY RESIDENT] much stronger predictors than [YOUNG] and [MALE] .
- [OLD] seems to be a stronger predictor of non-use than [YOUNG] is a predictor of use.

Preference



**Use in terms of
pairwise interactions
between variables
(cross-tabulations)**

Interaction of density and spread

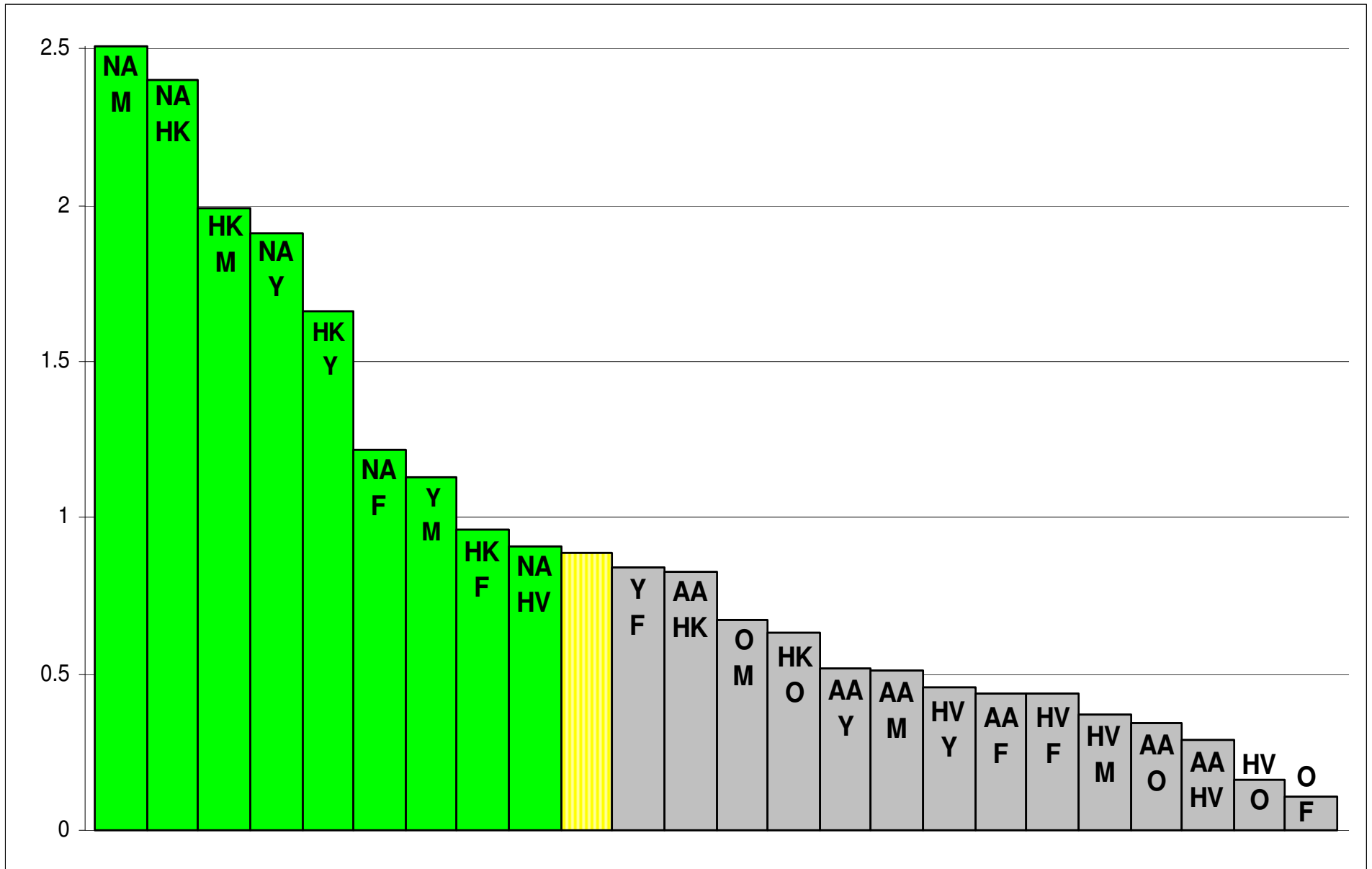


Most frequent values in combinations with above-average density and spread

Group	Freq. in pairs	Possible combinations	% of possible combinations
Non-Anglo	4	5*	80%
Hackney	4	6	67%
Male	3	6	50%
Young	3	6	50%
Female	2	6	33%

* There are no old non-Anglo speakers in the sample.

Pairwise preference



Comments and conclusions

- Ethnicity and residence stronger predictors than age and sex.
- Sequence of predictive utility:
ETHNICITY > RESIDENCE > AGE > SEX
- Network score not a dependable predictor overall.
- No discernible phonological patterns.
- The analysis assumed that all corpus speakers produced the same or similar number of words.
- The analysis in terms of number of words per speaker may reveal different patterns of use.